Beyond Autonomy: The Self and Life of Social Agents

Blue Sky Ideas Track

Budhitama Subagdja
ST Engineering-NTU Corporate Laboratory
School of Electrical and Electronic Engineering
Nanyang Technological University
Singapore
budhitama@ntu.edu.sg

Ah-Hwee Tan
School of Computer Science and Engineering
Nanyang Technological University
Singapore
asahtan@ntu.edu.sg

ABSTRACT

Agents have gained popularity nowadays as virtual assistants and companions of their human users supporting daily activities in many aspects of personal life. Designed to be sociable, an agent engages its user(s) to communicate and even develop friendships. Rather than just as a lifeless toy, it is supposed to be perceived as an individual with its own personality, experiences, and social life. In this paper, we seek to highlight self-hood as another dimension that characterizes an agent. Besides levels of autonomy and reasoning, an agent can be defined based on its capacity to process and reflect on its own self as an individual that possesses identity, embodiment, mind (mental), social relationship with others, and experiences comprising memories about the past and future prospects. We argue that this self-awareness is necessary for a companion agent to engage seamlessly with people as a real actual individual. Some existing implementations and models from preliminary works on agent's self-awareness illustrate the feasibility and challenges to realize this concept. Beyond assistance and companionship, we also envisage that this model of self is applicable to other types of autonomous application and system involving extensive interaction with people potentially tackling moral and ethical issues.

KEYWORDS

Self-Awareness; Autonomy; Social Cognition

ACM Reference Format:
Budhitama Subagdja and Ah-Hwee Tan. 2019. Beyond Autonomy: The Self and Life of Social Agents. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019*, IFAAMAS, 5 pages.

1 INTRODUCTION

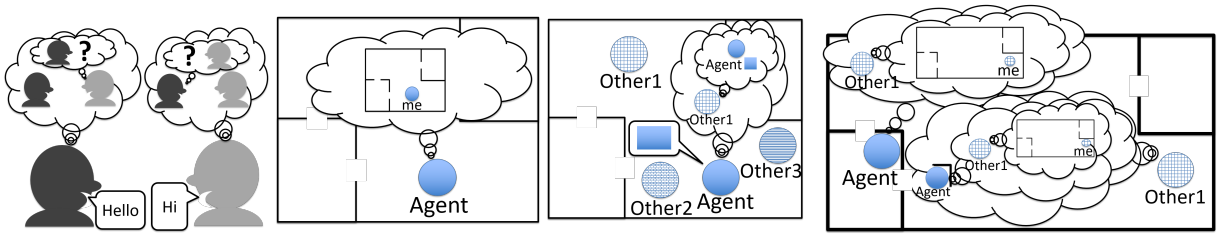
Autonomous agents have been widely adopted as virtual assistants or artificial companions to support their human user(s) in daily activities and personal life [6, 33]. Beyond performing repetitive or routine tasks, agents may engage users to communicate mostly through natural conversations and even develop relationships [34]. They are commonly built in various forms ranging from active applications resided in mobile devices to embodied human-like robots. These products of social agents are meant to be perceived and believed as if they are present as real individuals. People have started to talk with the agents as if they are real humans with mind and personality of their own [44] and even able to mutually teach and learn together [14].

However, instead of making the responses genuinely from the agents' own feelings or thoughts, they are merely expressed just to create the impression of their person-like social presence as a metaphor. In this case, there is no need for an agent to have its own view towards itself nor to maintain a model of its own self [15, 19] to fulfill this requirement. The agent's responses can be made pre-programmed or pre-configured with no underlying feelings nor thoughts. This non-realistic social competence limits the functionality of the agents. The impression that the agent is really present may happen only momentarily and cannot last long enough to develop close interpersonal relationships.

On the other hand, reflection on one's own self and reasoning about how one relates to another are essential in human-to-human communication. Communication is used by all participants for learning, influencing, assisting, socially relating, and playing from and to one another [37]. In *transactional model* of communication [3, 5], it is viewed that the participants act continuously and simultaneously as both speaker and listener while they develop shared meaning and mutual understanding according to their own past experiences, attitudes, self-images, or expectations. Knowing oneself becomes crucial since turn-taking in conversation [17] and meaning construction [8, 31] constitutes some reflection on one's own thought and anticipation on what and how the other thinks about oneself (Figure 1a). Furthermore, how deep the participants are willing to reveal information about internal states or private concepts of themselves to each other indicates the closeness of their interpersonal relationship [1]. *Self-disclosure* as the voluntary sharing of personal experiences, feelings, and other private inner states of mind becomes significant for ones to develop interpersonal relationships [1, 45].

Consequently, self-awareness which is defined as the capacity to put oneself (or the self of its own) as the focus of attention [16, 29, 32, 36] including understanding how others see oneself [18], is also essential for a social agent to develop trust and close acquaintance with others. This capacity allows one to continually learn about one's own social roles, relationships, and environment while in turn update the model and concept about oneself accordingly. It has been indicated that exhibiting social abilities like self-control or awareness of own mistakes of a companion robot can improve the sense of social presence and user acceptance [22]. Following this matter, we view that the current limitation in making the agent socially realistic is due to the lack of its incorporation and ascription with self-hood.

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.



(a) Anticipation in Conversation **(b)** Localization and mapping **(c)** Telling Identity **(d)** Playing hide and seek

Figure 1: Different Scenarios on the use of knowledge about self and self-awareness.

In this paper, we envisage the ascriptions of self and self-awareness for social robots and conversational agents to enable or to improve their social behavior, communication, and interpersonal relationship with humans. Besides the motivation to come up with a general framework of self-awareness for social agents, the conceptualization may also provide new insights on how self and self-awareness can be realized as parts of artificial general intelligence in the social context. In the rest of the paper, we discuss how self-hood is conceptualized as a property of a computational agent. This may add another dimension to classify or characterize an agent beyond their reasoning capacity or levels of autonomy. Based on the current existing preliminary works on self-awareness on robots and virtual agents, we discuss the feasibility of realizing the concept. We also put forward the gaps and challenges that still need to be overcome by the current technology of agents and AI in general. We also envisage that this concept of self-awareness can be applicable to other domains in general beyond companionship that may need extensive interaction with people or social environment.

2 SELF IN ROBOTS AND AGENTS

Despite the long-term history of ideas and conceptualization of self and self-awareness in many fields of study, a little attention is still given to the capacity to reflect on oneself in social robots or agents. Even though abundant examples of the robot's self-awareness have also been narrated in science fiction stories (droids and robots in Star Wars trilogy and series, Terminator, Star Trek TNG to name a few), the concept of self has not been much considered as parts of the design of a real social robot or a conversational agent.

In robotics, works in self-awareness focus on how a robot can learn and discover its own motion dynamic by testing and experiencing its own movements and actions [9, 21, 40]. This includes recognizing itself in the mirror [21, 40], or control and adapt its own movement robustly despite impairments [9]. Besides the dynamics of embodiment, continuously identifying self-properties within an environment like Simultaneous Localization-and-Mapping (SLAM) for robots navigation can also be considered as a particular form of self-awareness [27] (e.g the scenario in Figure 1b). Robots with this model of self are made to focus on modeling the embodiment, physical characteristics, and how they fit into their environment.

Similarly, how self-awareness occurs and relates to reasoning, planning, learning, and other cognitive and mental faculties have been computationally modeled in [11, 38, 39, 43]. These models view that self-awareness is a kind of meta-level cognition in which the system or the program forms and makes use of a representation about its own internal process in achieving its domain task objectives. One model emphasizes the role of autobiographical memory that the self-aware system represents multiple instances of itself in its mind as self-projections to different contexts and time from past experiences stored in memory [38].

Kuonev et al. [28] generalize this concept of self-awareness to characterize a computing system. Although the conceptualization can be considered more comprehensive and general than the meta-cognitive models above, it is still restricted to the internal aspects of the system itself isolated from its surrounding social world. A similar concept is made less restrictive by including *self-expression* as a part of the framework for engineering self-aware systems [30]. However, it is still limited to be just about displaying or presenting knowledge about self developed so far to other systems or agents. Bringsjord [10] has demonstrated a high level logical reasoning over self knowledge of a humanoid robot that recursively comprises others' self knowledge but only for making it to respond to a particular question of a very specific classical problem in epistemic logic.

In this paper, the envisioned model of self and self-awareness is considered should be as comprehensive as possible to sufficiently allow a second-person (or *n*th-person) to characterize and identify oneself (the first person). By empathizing one another reciprocally and dynamically in the so called *affective loop* [25, 26], both parties in the interaction will ultimately achieve self-awareness [2]. Subagdja and Tan [42] have proposed a model of self and self-awareness for a conversational humanoid robot similar to the envisioned concept in this paper. The robot represents multiple aspects of self in its mind with a complex nested structure of inter-relatedness of selves among parties to handle simple interaction and conversation. This last mentioned model is extended and elaborated in this paper.

3 KNOW THYSELF

In this paper, we adopt the common view in neuropsychology and brain science that consider self as an emerging phenomena from the interaction among different mental faculties in the brain [12, 24]. This implies that self-awareness constitutes constructive mental processes wherein information about oneself either as the subject, object, or every aspect that belong to the individual is being attended in one's own mind.

This conforms with evidence in neuropsychology that awareness about the presence of limbs in one's body is mentally made up [35] by the individual in order to make sense one's own existence [4, 7, 12, 24].

3.1 Levels of Self-Awareness

The constructive process of self-awareness in a social agent, in this case, can be formulated to constitute the creation of instances of information or knowledge that refer to something. This mental construct can be called *awareness* which represents someone or something in the world, an imaginary object, an abstract thought, or a snapshot of memory flashback. For instance, when an agent i is aware of or that p, it means that p is something in i's mind or mental state. Figure 2 illustrate this as *basic awareness*. However, self-awareness occurs only if p contains a description or reference to i itself as the agent that possesses the awareness. Consequently, awareness about something still does not count as self-awareness if no reference to the agent's own self is present even though it is expressed in a complex nested structure (e.g i aware that j aware that k ... where $i \neq j \neq k \neq \ldots$). In this paper, awareness can take different forms as an aspect of mind or mental state. Any type of intentional modality in *theory of mind* can be considered as a kind of awareness like *want*, *observe*, *recall*, *think*, or *imagine*.

There are two kinds of self-awareness in terms of the level of the agent's point of view:

- ***first-agent*** self-awareness in which the agent's self is seen from the first-person view or as the subject of the awareness. For example, Figure 2 shows that agent (robot) i is aware that i itself is aware of p (e.g think about, remember, or see p). This is also called *subjective* self-awareness [42];
- ***nth-agent*** self-awareness in which the agent's self is seen from the second or nth person view or as the object of the awareness of another agent. For example, Figure 2 shows that agent i can be aware that p is aware of i. This can also be called *objective* self-awareness [42].

As a constructive process, awareness is created dynamically inside one's mind triggered or initiated by perception or any kind of thought about something (someone). In this case, self-awareness may have a complex nested structure representation about anything inside one's or the others' mind. The distinction of this envisioned concept with the other related models described above is that the representation of self may consist of complex intertwining relations among different individuals which may be projected into different contexts and time though they must have some references to oneself (*complex self-awareness* in Figure 2).

3.2 Aspects of Self

Besides the complex nested structure of representation, self should also be tightly related to the agent's social world [4, 7, 18]. Besides the levels of self-awareness, the types or aspects of information to characterize a social agent are necessary to define. These aspects can be envisaged as follows.

3.2.1 ***Identity.*** This aspect is about the information that can be used to characterize the agent as a unique individual or as a distinct entity from others. It may include categorical information or classification of oneself to identify the individual (e.g type, race, color, looks). Identity is the main information necessary for self to exist. From the subjective self-awareness, the agent should know its own identity information like name, registry number, race, or age. From the objective or nth-agent view, the scenario in Figure 1c exemplifies the revelation of the agent's identity information (color) to another, expecting the other identifies and ascribes it to the agent as its identity. A self-aware agent should develop and maintain its identity over its lifetime or its overall period of use. This identity information enables the agent to have its own social life distinct from the others and to maintain a long-term interpersonal relationship with humans.

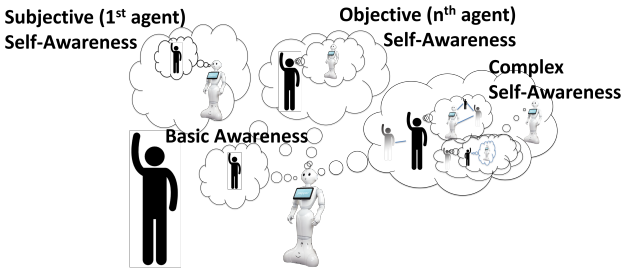


Figure 2: Different levels of awareness and self-awareness.

3.2.2 ***Embodiment.*** Embodiment is about the characterization of the embodiment, situation, and environment wherein the agent is present. This includes for example, knowing the bodily structure or limbs, pose, current location (relative to the environment), abilities to change the state of environment and so on. A simple example is the first-agent self awareness in locating the position of the agent's self with respect to the layout of the environment (Scenario in Figure 1c). Another example is nth-agent self awareness in playing hide and seek (Figure 1d). The last example shows that the agent can project its embodiment model to the other's mind expecting that the other the agent is not visible or not present. However, the other agent may also anticipate that the first agent could think in that way regarding the other's own awareness when it is located in a particular place (e.g in the corner of a room).

3.2.3 ***Mind.*** Mind aspect is about the characterization of an individual based on one's mental state. This includes the kind of reasoning, feelings, affective state, or mental operation that is present in one's mind at one moment. The agent may reflect on its own intention, desire, observation, what it recalls from memory, or what it has imagined as different forms of awareness. As mentioned previously, awareness can take many forms like desire, thought, imagination, perception, or feeling (emotion). With the objective kind of self-awareness it is possible to express intentional state in terms of the self aspects. This mind aspect enables the self representation to have its deep nested structure.

3.2.4 ***Relationship.*** Relationship aspect is about the characterization of an individual based on how it socially relates with other individuals. The subjective self awareness of the agent's relationship allows it to know and identify its friends, foes, owners, hosts, whom does it like (or dislike), and others socially related. In the nth-agent view, the agent becomes aware of whom or what others think or feel about the agent's self. This social aspect of self allows the agent to have a social goal like making friends. Together with the mind aspect of information, the agent may develop or emulate

a complex social emotion like shame, pride, or guilty which can only be initiated by the thought regarding self in the other's mind.

3.2.5 ***Memory.*** Memory aspect is about the characterization of an individual based on what one has experienced or what one thinks will unfold in the future based on experiences so far. This aspect constitutes memory of the agent, remembrance, and foresight about the future prospect beyond momentary awareness and thought. In this aspect, one requires an autobiographical memory to maintain multiple version of oneself in different time frame as a projection of one in the past and another version of self projected to the future [38]. In this case, the agent can represent every aspect of self as above in temporally extended manner.

3.3 Beyond Autonomy

Agents are commonly characterized based on their autonomy [20, 46]. However, this characterization has still a broad spectrum. It is still challenging to determine the exact boundary between one type and the other level of autonomy. An autonomous agent may be equipped with a sophisticated machine learning algorithm to automatically acquire knowledge from its experiences and use it to accomplish a certain task. However, another agent that behave reactively with simple pre-programmed rules to achieve the same task can still be considered as autonomous. Self-hood can be an alternative for characterizing an agent to overcome this vagueness. Which level of self representation can be processed and what aspects of information about self are covered give clear boundaries to distinguish one type of agent from another since it only depends on the kind of information it is used to represent the properties of self. Based on this characterization, an agent may someday no longer be defined as an autonomous entity but seen as one that represents oneself on its self knowledge and reason based on those representation to some extent.

4 CHALLENGES

Despite successful implementations on continuous self modeling of physical and dynamics of embodiment in robotics [9, 21, 40], there are still challenges that must be tackled to make this conceptualization of self and self-awareness practical. The first challenge is the lack of formal study and precise definition of their aspects. Those aspects and types of information described above are still intuitive and still based on relatively vague categorization. More thorough studies are still needed to investigate the proper aspects and formalize them in the model. The second challenge is related to the computational complexity to handle the complex representation of self as presented above. Processing the nested intertwining structure of different awareness representation can be intractable. There is no clear guiding principle to limit the depth of nth-agent self-awareness representation. Interestingly, Subagdja and Tan [42] proposes a model of working memory for self-awareness that can handle the nested structure of self-awareness similar to the model presented here. The nested structure of self-awareness is built using a neural network architecture which can self-organize the levels generated in the nested structure of awareness. This avoids infinite generation of awareness levels and can demonstrate that the mind aspect of self-awareness is potentially practical. Another computational challenge is to handle the long lifetime application of autobiographical memory wherein the system must store all relevant experiences including the awareness structure as described above. This kind of system may require very large storage capacity for each individual agent. Each of the challenges mentioned above is still an open research question and engineering problem. Another study necessary to pursue is to look at the user or human acceptance and the believability of social presence of this kind of social agents with self awareness.

5 TACKLING MORAL AND ETHICS

Scenarios that exemplify aspects of self shown above has indicated that the envisioned concept of self is applicable to domains wherein human involvements are not necessary. For example, a multi-agent system dealing with a critical mission like agent-to-agent collaboration to localize and to map the environment in a search and rescue mission or security surveillance may require the agents to anticipate what the others think about themselves and act upon it (see hide and seek scenario in Figure 1c). The objective self-awareness in anticipating others may also allow the agents to learn and represent some norms and moral codes on the fly. For example, a self-driving car embedded with the self-awareness may be able to learn the norms or moral code of driving on the road directly based on its observation of how other drivers behave and think about the car itself. In a certain case, the car may initiate a non-verbal communication with a human driver (e.g by "honk the horn").

A possible way to realize this kind of self-awareness agent is by embedding it to an existing component or service so that it becomes aware of itself in terms of the aspects described in this paper and able to reveal its own internal states and potential collaboration to its human developer or user (e.g teachable components [41]). Another interesting application of this embeddable self-awareness is for realizing explainable artificial intelligence (XAI) [13, 23] wherein the internal states and processes of an AI system can be communicated transparently.

6 CONCLUSION

In this paper we have presented the concept of self for an agent supporting their capacity to be sociable and to engage its user(s) to communicate and develop relationships. The agent can truthfully be perceived as an individual with its own identity, experiences, and social life. We have highlighted the concept of self-hood as another dimension that characterizes an agent besides levels of autonomy and reasoning. The concept allows an agent to be defined based on its capacity to process and reflect on its own self as an individual that possesses identity, embodiment, mind (mental), social relationship with others, and experiences comprising memories about the past and future prospects. Models and design of virtual assistants and social robots from some existing preliminary works on agent's self-awareness have indicated the feasibility and challenges to realize this concept. we also have envisaged that this model of self is applicable to other types of autonomous application and system involving collaboration and extensive interaction with people.

ACKNOWLEDGMENTS

The research was partially supported by the ST Engineering - NTU Corporate Lab through the NRF corporate lab@university scheme.

REFERENCES

[1] Irwin Altman and Dalmas A. Taylor. 1973. *Social Penetration: the development of interpersonal relationships.* Holt, New York.

[2] Minoru Asada. 2015. Towards Artificial Empathy: How Can Artificial Empathy Follow the Developmental Pathway of Natural Empathy? *International Journal of Social Robotics* 7, 1 (2015), 19–33.

[3] Dean C. Barnlund. 1970. A Transactional Model of Communication. In *Foundations of Communication Theory*, Kenneth K. Sereno and David C. Mortensen (Eds.). Harper, New York, 83–102.

[4] Lissa Feldman Barrett. 2017. *How Emotions Are Made: The Secret Life of the Brain.* Houghton MIfflin Harcourt, New York.

[5] Samuel L. Becker. 1999. Rhetorical studies for the contemporary world. *Communication Studies* 50, 1 (1999), 28–44.

[6] Susanne Biundo, Daniel Höller, Bernd Schattenberg, and Pascal Bercher. 2016. Companion-Technology: An Overview. *Künstliche Intelligenz* 30, 1 (2016), 11–20.

[7] Sarah-Jayne Blakemore. 2018. *Inventing Ourselves: The Secret Life of the Teenage Brain.* PublicAffairs, New York.

[8] Herber Blumer. 1969. *Symbolic Interactionism: Perspective and Method.* Prentice-Hall, Englewood-Cliffs.

[9] Josh Bongard, Victor Zykov, and Hod Lipson. 2006. Resilient Machines Through Continuous Self-Modeling. *Science* 314, 5802 (2006), 1118–1121.

[10] Selmer Bringsjord, John Licato, Naveen Sundar Govindarajulu, Rikhiya Ghosh, and Atriya Sen. 2015. Real robots that pass human tests of self-consciousness. In *Proceedings of the Twenty-Fourth IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2015).* 498–504.

[11] Michael T. Cox. 2007. Perpetual Self-Aware Cognitive Agents. *AI Magazine* 28, 1 (2007), 32–45.

[12] Antonio Damasio. 2010. *Self comes to mind: Constructing the conscious brain.* Pantheon/Random House, New York.

[13] Filip Karlo Došilović, Mario Brčić, and Nikica Hlupić. 2018. Explainable artificial intelligence: A survey. In *Proceedings of the 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO 2018).* 210–215.

[14] Stefania Druga, Randi Williams, Cynthia Breazeal, and Mitchel Resnick. 2017. "Hey Google is it OK if I eat you?": Initial Explorations in Child-Agent Interaction. In *Proceedings of the 2017 Conference on Interaction Design and Children (IDC'17).* 595–600.

[15] Paul Dumouchel and Luisa Damiano. 2017. *Living with robots.* Harvard University Press, Cambridge.

[16] Shelley Duval and Robert A. Wicklund. 1972. *A Theory of Objective Self Awareness.* Academic Press, New York.

[17] N. J. Enfield. 2017. *How We Talk: The inner workings of conversation.* Basic Books, New York.

[18] Tasha Eurich. 2017. *Insight: Why We're Not as Self-Aware as We Think, and How Seeing Ourselves Clearly Help Us Succeed at Work an in Life.* Crown Business, New York.

[19] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. 2003. A Survey of Socially Interactive Robots: Concepts, Design, and Applications. *Robotics and Autonomous Systems* 42, 3–4 (2003), 143–166.

[20] Stan Franklin and Art Graesser. 1997. Is It an agent, or just a program?: A taxonomy for autonomous agents. In *Intelligent Agents III Agent Theories, Architectures, and Languages ATAL 1996*, P. J. Müller, Michael J. Wooldridge, and Nicholas R. Jennings (Eds.). LNCS, Vol. 1193. Springer.

[21] Justin W. Hart and Brian Scasselati. 2012. Mirror Perspective-Taking with a Humanoid Robot. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence (AAAI 2012).* 1990–1996.

[22] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. 2008. The Influence of Social Presence on Acceptance of a Companion Robot by Older People. *Journal of Physical Agents* 2, 2 (2008), 33–40.

[23] Robert R. Hoffman, Gary Klein, and Shane T. Mueller. 2018. Explaining Explanation For "Explainable AI". *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 62, 1 (2018), 197–201.

[24] Bruce Hood. 2012. *The Self Illusion: How the Social Creates Identity.* Oxford University Press, New York.

[25] Kristina Höök. 2008. Affective Loop Experiences – What Are They?. In *Proceedings of Third International Coference on Persuasive Technology (PERSUASIVE 2008).* 1–12.

[26] Kristina Höök. 2009. Affective loop experiences: designing for interactional embodiment. *Philosophical transactions of Royal Society of London. Series B, Biological sciences* 364, 1535 (2009), 3585–3595.

[27] Yoshiteru Ishida. 2015. A Note on Continuous Self-Identification as Self-Awareness: An Example of Robot Navigation. *Procedia Computer Science* 60, 2015 (2015), 1865–1874.

[28] Samuel Kounev, Jeffrey O Kephart, Aleksandar Milenkoski, and Xiaoyun Zhu (Eds.). 2017. *Self-Aware Computing Systems.* Springer, Cham.

[29] Laure Legrain, Axel Cleeremans, and Arnaud Destrebecqz. 2011. Distingushing three levels in explicit self-awareness. *Consciousness and Cognition* 20, 3 (2011), 578–585.

[30] Peter R. Lewis, Marco Platzner, Bernhard Rinner, Jim Tørresen, and Xin Yao (Eds.). 2016. *Self-Aware Computing Systems: An Engineering Approach.* Springer.

[31] George Herber Mead. 2015. *Mind, Self, and Society: the definitive edition.* University of Chicago Press.

[32] Alain Morin. 2006. Levels of consciousness and self-awareness: A comparison and integration of various neurocognitive views. *Consciousness and Cognition* 15, 2 (2006), 358–371.

[33] Matt O'Brien. 2018. Robots are getting more social. Are humans ready? *Associated Press* August 9, 2018 (2018). https://apnews.com/abf8fce4539d405c831a25a46b13a40d

[34] Adam Piore. 2014. Friend For Life. *Popular Science* 285, 5 (2014), 38–44, 83–84.

[35] Vilayanur S. Ramachandran. 2011. *The Tell-Tale Brain: A Neuroscientist's Quest for What Makes Us Human.* W.W. Norton & Company, New York.

[36] Philippe Rochat. 2003. Five levels of self-awareness as they unfold early in life. *Consciousness and Cognition* 12, 2003 (2003), 717–731.

[37] Rebecca B. Rubin, Carlos Fernàndez, and Roberto Hernandez-Sampieri. 1992. A cross-cultural examination of interpersonal communication motives in Mexico and The United States. *International of Journal of Intercultural Relations* 16, 2 (1992), 145–157.

[38] Alexei V. Samsonovich and Lynn Nadel. 2005. Fundamental Principles and Mechanisms of the Conscious Self. *Cortex* 41, 5 (2005), 669–689.

[39] Aaron Sloman and Ron Chrisley. 2003. Virtual Machines and Consciousness. *Journal of Consciousness Studies* 10, 4–5 (2003), 133–172.

[40] Alexander Stoytchev. 2011. Self-detection in robots: a method based on detecting temporal contingencies. *Robotica* 29, 1 (2011), 1–21.

[41] Budhitama Subagdja and Ah-Hwee Tan. 2016. Interactive Teachable Cognitive Agents: Smart Building Blocks for Multiagent Systems. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 46, 12 (2016), 1724–1735.

[42] Budhitama Subagdja and Ah-Hwee Tan. 2017. Towards a Brain Inspired Model of Self-Awareness for Sociable Agents. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17).* 4452–4458.

[43] Ron Sun. 2007. The importance of Cognitive Architecture: An Analysis based on CLARION. *Journal of Experimental and Theoretical Artificial Intelligence* 19, 2 (2007), 159–193.

[44] Daniel Terdiman. 2018. Here's How People Say Google Home and Alexa Impact Their Lives. *Fast Company* 01.05.18 (2018). https://www.fastcompany.com/40513721/heres-how-people-say-google-home-and-alexa-impact-their-lives

[45] Betsy E. Tolstedt and Joseph P. Stokes. 1984. Self-disclosure, intimacy, and the depenetration process. *Journal of Personality and Social Psychology* 46, 1 (1984), 84–90.

[46] Michael Wooldridge. 2009. *An Introduction to MultiAgent Systems* (second edition ed.). John Wiley & Sons Ltd, Chichester.